AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2015WR017044

Key Points:
- We select the most valuable information for water systems operation
- We assess the operational and economic value of this information
- We describe how to directly embed exogeneous information into the operating policies

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Giuliani,
matteo.giuliani@polimi.it

Citation:
Giuliani, M., F. Pianosi, and A. Castelletti (2015), Making the most of data: An information selection and assessment framework to improve water systems operations, *Water Resour. Res.*, *51*, 9073–9093, doi:10.1002/2015WR017044.

Received 9 FEB 2015
Accepted 25 OCT 2015
Accepted article online 30 OCT 2015
Published online 19 NOV 2015

Making the most of data: An information selection and assessment framework to improve water systems operations

M. Giuliani[1], F. Pianosi[2], and A. Castelletti[1,3]

[1]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy, [2]Department of Civil Engineering, University of Bristol, Bristol, UK, [3]Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland

Abstract Advances in Environmental monitoring systems are making a wide range of data available at increasingly higher temporal and spatial resolution. This creates an opportunity to enhance real-time understanding of water systems conditions and to improve prediction of their future evolution, ultimately increasing our ability to make better decisions. Yet, many water systems are still operated using very simple information systems, typically based on simple statistical analysis and the operator's experience. In this work, we propose a framework to automatically select the most valuable information to inform water systems operations supported by quantitative metrics to operationally and economically assess the value of this information. The Hoa Binh reservoir in Vietnam is used to demonstrate the proposed framework in a multiobjective context, accounting for hydropower production and flood control. First, we quantify the expected value of perfect information, meaning the potential space for improvement under the assumption of exact knowledge of the future system conditions. Second, we automatically select the most valuable information that could be actually used to improve the Hoa Binh operations. Finally, we assess the economic value of sample information on the basis of the resulting policy performance. Results show that our framework successfully select information to enhance the performance of the operating policies with respect to both the competing objectives, attaining a 40% improvement close to the target trade-off selected as potentially good compromise between hydropower production and flood control.

1. Introduction

In a rapidly changing context, where climate change and growing populations are straining freshwater availability worldwide, using existing infrastructures more efficiently, rather than planning new ones, becomes key to balance competing objectives and performance uncertainties, while minimizing investments and financial risk [*Gleick and Palaniappan*, 2010]. Many large storage projects worldwide have had their operations designed in prior decades [*U.S. Army Corps of Engineers*, 1977; *Loucks and Sigvaldason*, 1982], with operating rules often conditioned on very simple information systems, such as inflow in the current time period and previous release [*Hejazi et al*., 2008]. Under changing hydroclimatic and socioeconomic forcing, the reluctance to adapt reservoirs operations to the new conditions [e.g., *Sheer*, 2010; *Fernandez et al*., 2013; *Giuliani et al*., 2014a] has led many large-scale dams to fail in producing the level of benefits that provided the economic justification for their development [e.g., *Stone*, 2011; *Ziv et al*., 2012; *Ansar et al*., 2014].

The unprecedented "torrent of information" [*The Economist Editorial*, 2011] that is becoming increasingly available to water system operators from pervasive sensor networks [e.g., *Hart and Martinez*, 2006], remote sensing [e.g., *Butler*, 2007], cyberinfrastructure [e.g., *Minsker et al*., 2006], and crowdsourcing [e.g., *Fraternali et al*., 2012], combined with the advances in data analytics and optimization techniques [e.g., *Maier et al*., 2014], creates an opportunity for improving water systems operations in novel, unconventional ways, and with minor investments. However, while this information might be useful to improve our understanding and prediction of environmental processes, it also introduces observational errors and estimation biases that challenge its optimal use.

In this paper, we propose a novel information selection and assessment (ISA) framework to support the efficient use of observational data and improve water systems operations. The analysis of the value of information explicitly quantifies potential losses from limited knowledge and uncertainty of current and future system conditions, and identifies the "best" information as the one leading to the greatest expected benefits [*Yokota and Thompson*, 2004].

Our ISA framework is composed by three interconnected steps. We first compute the expected value of perfect information (EVPI), meaning the value of completely eliminating uncertainty from the decision-making process [*Yokota and Thompson*, 2004]. In fact, a water resources system, such as a water supply distribution system or a network of reservoirs, can be underperforming for a number of reasons, including structural limitations in the system's infrastructure [e.g., *Castelletti et al*., 2012] or in the institutional settings [e.g., *Madani and Lund*, 2012; *Giuliani and Castelletti*, 2013; *Anghileri et al*., 2013], as well as the lack or inaccuracy of the information (e.g., flow observations and/or forecasts) used to inform operational decisions. Our framework focus on the latter and quantify the benefit that would be obtained if perfect information were available, considering the multiple, potentially conflicting, operating objectives of the system under study as well as their associated monetary value. The EVPI is estimated by the difference between the system performance that could be obtained if perfect information were available and optimally used (upper bound solution) and the system performance in an uninformed baseline alternative (lower bound solution) [*Delquié*, 2008]. A positive EVPI indicates a potential benefit in collecting more information [*Khader et al*., 2013] and quantifies the maximum price that a decision maker would be willing to pay to obtain such information.

Estimating the EVPI is useful to increase the system understanding and to support decision making in long-term planning, for instance to prioritize investments in physical infrastructure versus investments in information infrastructure. However, in real-time, uncertainty cannot be completely eliminated and perfect information can remain unavailable. We therefore compute another measure, the expected value of sample information (EVSI), which assesses the value of using the information actually available to the system operator when operational decisions are taken. For instance, for a reservoir system, perfect information would include exact knowledge of future reservoir inflows, while sample information would include some other observable quantities, such as current snowpack depth or upstream flows, which might be used to anticipate future inflows. The EVSI provides the value of reducing—but not eliminating—uncertainty [*Yokota and Thompson*, 2004]. Determining the EVSI requires exploring two main questions: How can we select, among the available information sources, the ones that more valuably contribute to the system operations? How can the operating policy be re-designed to cater for the selected information? Addressing these two challenges will constitute the second and third step of our framework.

In the literature, numerous studies have focused on these questions by incorporating basic hydrologic information, selected on the basis of operators' experience, in a dynamic programming framework adopted for the design of water reservoirs operations. Common choices have been the observations of previous period's inflows [e.g., *Bras et al*., 1983; *Tejada-Guibert et al*., 1995], simplified models of other hydrologic variables [e.g., *Côté et al*., 2011; *Desreumaux et al*., 2014], or streamflow forecasts [e.g., *Stedinger et al*., 1984; *Karamouz and Vasiliadis*, 1992; *Kim and Palmer*, 1997; *Faber and Stedinger*, 2001; *Maurer and Lettenmaier*, 2004; *Voisin et al*., 2006; *Shukla et al*., 2012; *Oludhe et al*., 2013; *Li et al*., 2014; *Zhao et al*., 2014]. As for the latter, the use of reliable inflow forecasts is beneficial under most situations. Yet, its real value is often problem specific and depends upon the system's dominant dynamics and the objectives considered [*You and Cai*, 2008; *Graham and Georgakakos*, 2010]. For example, using short-term inflow forecasts generally improves reservoir operations for flood control, [e.g., *Castelletti et al*., 2008a; *Pianosi and Soncini-Sessa*, 2009]. However, this improvement increases linearly with hydrologic uncertainty and decreases logarithmically with reservoir size [*Hejazi et al*., 2008]. For other objectives, for instance water supply, medium and long-term streamflow forecasts would be needed. Yet, while some works [e.g., *Hamlet and Lettenmaier*, 1999; *Sharma*, 2000; *Block and Goddard*, 2012; *Zhao and Zhao*, 2014] successfully extended the lead time of flow forecasts by using low-frequency climate phenomena (e.g., El Nino Southern Oscillation), in regions where the influence of these global phenomena is less intense (e.g., Europe or Africa), the quality of medium-long term streamflow forecasts is still relatively low [e.g., *Hansen et al*., 2011; *Sharma and Chowdhury*, 2011; *Alemu et al*., 2011] and their use for improving reservoir operations less beneficial. To overcome these limitations, in this paper we adopt an alternative approach and we investigate the potential for using observational data (e.g., past and current rainfall, flow, water levels, etc., possibly at different time lags) in the operations of water resources systems by directly conditioning the operating policies on observed variables without the intermediation of forecasting models (i.e., model-free).

Besides designing improved operating policies using the selected information, the third step of our ISA framework aims also at determining quantitative approaches to estimate the EVSI by contrasting the performance of these policies with the upper and lower bound solutions. Moreover, we illustrate an approach to estimate the economic value of using additional information. Since acquiring information requires money and time [*Sakalaki and Kazi*, 2007], the economic EVSI expresses the willingness to pay of a water authority, expressed as the maximum amount the water authority would pay to obtain the information required for the design and implementation of the informed operating policies [e.g., *Alberini et al.*, 2006].

In summary, this paper provides three main methodological contributions: (i) we propose a procedure to automatically select the most valuable information to inform water systems operations; (ii) we provide metrics to quantitative and economically assess the value of this information; (iii) we describe computational tools to directly embed this information into the operating policies in a model-free fashion, i.e., with no need of a forecasting model.

The paper is organized as follows: the ISA framework is described in the next section. Section 3 reviews some numerical methods that can be adopted to perform the different steps of the procedure. Section 4 introduces the Hoa Binh case study application used to demonstrate the potential for the proposed ISA framework. The Hoa Binh is a large reservoir on the Red River in Vietnam, mainly operated for hydropower production and flood control. This case study has been chosen for its medium-complexity, which means that it provides a realistic assessment of the proposed framework while also allowing for a physically meaningful interpretation of the results. Numerical results are reported in section 5, and final remarks and issues for further research are presented in the last section.

2. General Framework

2.1. Problem Statement

We consider a general multiobjective water reservoir operations problem of the following form:

$$\max_{\mathbf{u}_{[0,H-1]}} \mathbf{J}\left(\mathbf{x}_{[0,H]}, \mathbf{u}_{[0,H-1]}, \varepsilon_{[1,H]}\right) \tag{1a}$$

subject to

$$\mathbf{x}_{t+1}=f_t(\mathbf{x}_t, \mathbf{u}_t, \varepsilon_{t+1}) \qquad t=0, \ldots, H-1 \tag{1b}$$

$$\mathbf{x}_0, \varepsilon_{[1,H]} \text{ given} \tag{1c}$$

where

1. $\mathbf{u}_t$ is the vector of release decisions at each time step $t=0, \ldots, H-1$ over the evaluation horizon $[0,H]$.
2. $\mathbf{J}(\cdot)$ is the objective function vector, which accounts for K operating objectives (to be maximized) evaluated over the horizon H. As shown in equation (1a), the objective values depends on the sequence of release decisions $\mathbf{u}_{[0,H-1]}$, the trajectory of the external drivers $\varepsilon_{[1,H]}$ (e.g., reservoir inflows), and the resulting trajectory of the state vector $\mathbf{x}_{[0,H]}$ (e.g., reservoir storage). The model used for describing the external drivers $\varepsilon_{[1,H]}$ depends on the problem formulation as discussed in the next sections.
3. $f(\cdot)$ is the transition function of the system (e.g., reservoirs' water balance and flow-routing), whose recursive application allows the dynamic simulation of the system's evolution in time. In the adopted notation, the time subscript of a variable indicates the time instant when its value is deterministically known. The reservoir storage is measured at time t and thus is denoted as x_t, while inflow in the interval $[t,t+1)$ is denoted as ε_{t+1} because it can be known only at the end of the time interval.

The ISA framework is composed by the three main building blocks illustrated in Figure 1, which are described one by one in the following sections. Note that the ISA framework should be applied in an iterative fashion until the performance of the designed operating policy meets the level of performance required by the decision makers or when adding new information does not further improve the system performance.

2.2. Quantifying the Expected Value of Perfect Information

We define the Expected Value of Perfect Information (EVPI) as the performance improvement that could be achieved under the assumption of perfect foresight on the future at the moment when operational decisions must be taken. The optimal sequence of release decisions ($\mathbf{u}^{POP}_{[0,H-1]}$) that an ideal operator would have followed having perfect information on the future is obtained by solving Problem (1) with the trajectory $\varepsilon_{[1,H]}$ of the external drivers deterministically known over the entire evaluation horizon H. The assumption of

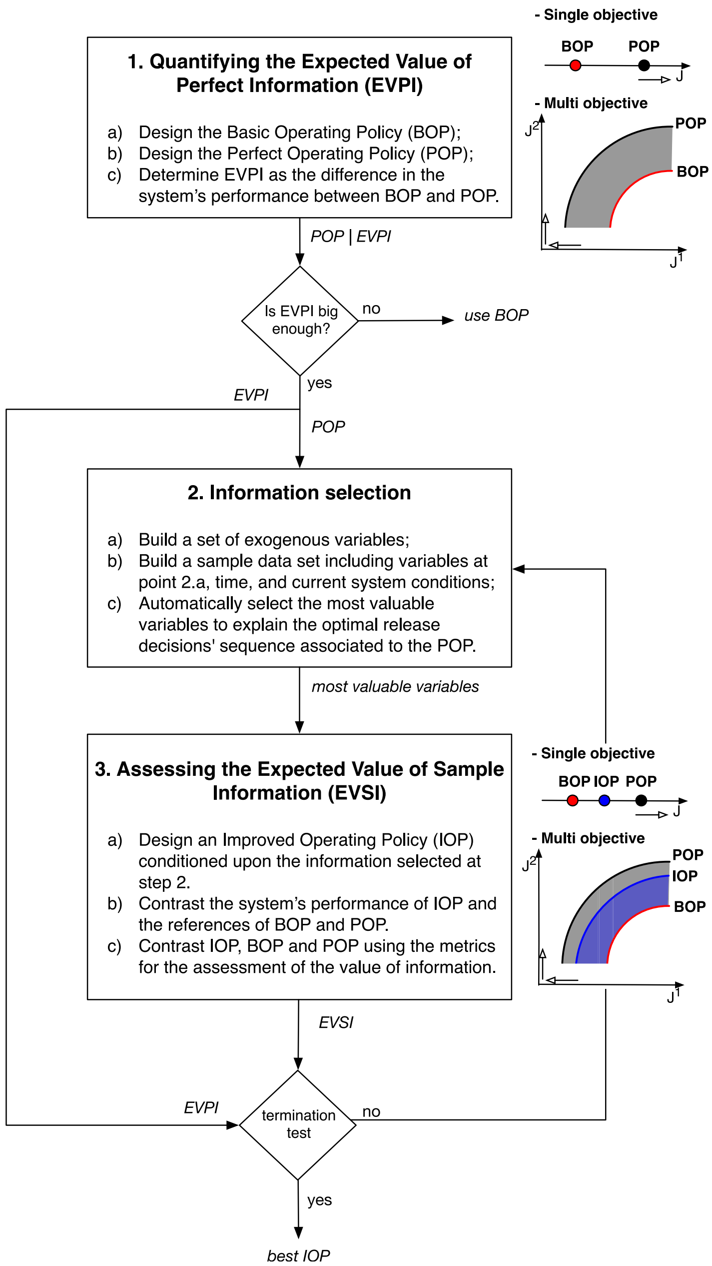

Figure 1. Schematization of the Information Selection and Assessment (ISA) framework.

perfect (deterministic) knowledge of the future external drivers implies that this solution is designed as an open loop operating policy. However, this sequence of optimal release decisions is conceptually equivalent to a closed loop operating policy conditioned on the current system conditions, represented by the time instant t and the state vector $\mathbf{x}_t$ [*Bertsekas*, 1976], along with perfect information on the future. In fact, the decision vector $\mathbf{u}_t$ at each time t is an implicit function of the current state vector $\mathbf{x}_t$ and of the sequence of future external drivers $\varepsilon_{[t+1,H]}$, assumed to be perfectly known at time t (Figure 2a). We call Perfect Operating Policy (POP) such a reconstructed optimal solution.

The POP performance (J^{POP}) is an absolute measure of the system performance, which is not completely meaningful as it depends on the characteristics of the system under study, for instance the reservoir capacity-inflow ratio. As a consequence, the EVPI has to be estimated by comparing J^{POP} with the value of the objective functions that could be obtained, over the same evaluation horizon, by a poorly informed operating policy (i.e., Basic Operating Policy—BOP) relying on a basic set of information. For example, we can consider the so called release plan [e.g., *Soncini-Sessa et al.*, 2007], which means an open loop operating policy that depends only on the day of the year (i.e., $\mathbf{u}_t = BOP(t)$) and no information on the future external drivers. In case of a single-objective problem, the EVPI is given by the difference $J^{POP} - J^{BOP}$ between the (scalar) performance of the POP and BOP. In the multiobjective case, the objective functions J^{POP} and J^{BOP} are vectors and the estimation of the EVPI is more complex. Quantitative metrics to estimate the EVPI from the comparison of vector objective functions are presented in section 3.4. In both cases, a positive and sufficiently large EVPI indicates a potentially significant benefit from using more information to improve the operations of the system, whereas a small EVPI suggests that improvement of the Basic Operating Policy would have limited effect.

2.3. Information Selection

When the EVPI is large, in order to close the gap between POP and BOP, we need to identify a set of sample information $\mathbf{I}_t$, observable at time t, that can act as an effective surrogate of the sequence of future external

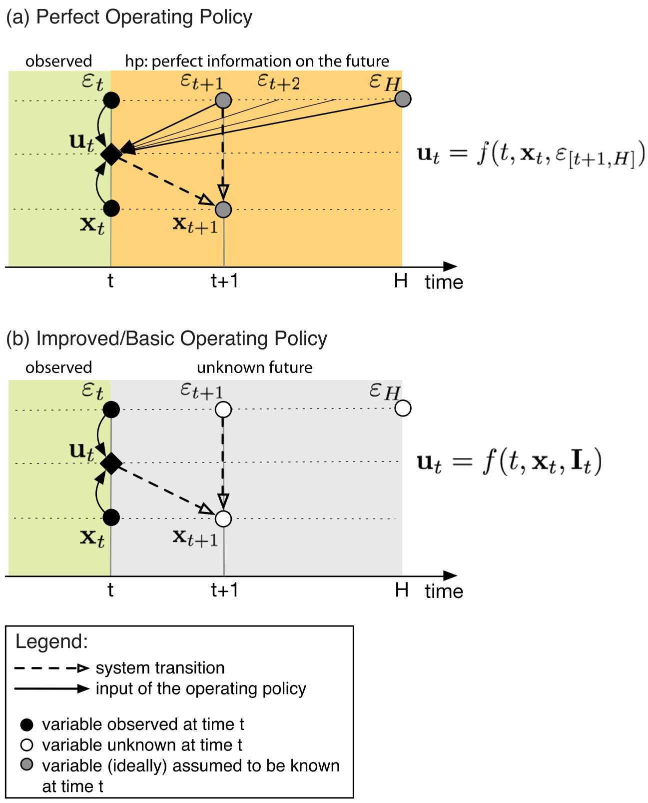

Figure 2. (a) Different use of information for making decisions under the Perfect Operating Policy, assuming perfect information on the future is available, and (b) under the Basic and Improved Operating Policies, assuming only the information actually available at the time instant of the decision is used.

drivers and characterizes as much accurately as possible the optimal sequence of release decisions $\mathbf{u}^{POP}_{[0,H-1]}$. In principle, streamflow forecasts can be considered in this step as candidate information. However, according to our model-free approach, we prefer to select information from observed variables only.

The set Ξ_t of candidate variables hence comprises any exogenous variable (i.e., variables that are observed but are not endogenous in the problem formulation and hence are not modeled) such as observed rainfall, flows, water levels, etc. at various locations within the system and, possibly, at different time lags or integrated over different periods. The first problem to address is how to efficiently select the smallest subset of variables $\mathbf{I}_t \subseteq \Xi_t$ carrying the most valuable information. Since the set of candidate exogenous variables Ξ_t can be rather vast, comprising redundant and collinear variables, a numerical procedure may help to tackle the problem.

To this end, we suggest to support this selection by adopting automatic methods capable of identifying the sub-set of variables $\mathbf{I}_t \subseteq \Xi_t$ that better explain the optimal sequence of release decisions $\mathbf{u}^{POP}_{[0,H-1]}$. This sequence characterizes the optimal operations of the system in the ideal case of perfect knowledge of the future conditions. Several methods can be adopted to solve such information selection problem, such as cross-correlation or mutual information analysis, or more complex input variable selection techniques. Technical details about these methods are further discussed in section 3.2.

2.4. Assessing the Expected Value of Sample Information

Once we have selected the best surrogate $\mathbf{I}_t$ of the future external drivers, the next step is to design the Improved Operating Policy (IOP) that uses this surrogate information to inform operational decisions. The IOP is a closed loop operating policy that provides the decision vector as a function of the information available at time *t*, i.e., $\mathbf{u}_t = IOP(t, \mathbf{x}_t, \mathbf{I}_t)$ (Figure 2b).

Two alternative approaches are available to design the IOP: (i) to identify a dynamic model describing each component of $\mathbf{I}_t$ and use the states of these models to condition the operating policies within a dynamic programming framework [e.g., *Tejada-Guibert et al*., 1995; *Desreumaux et al*., 2014]; (ii) to adopt approximate dynamic programming methods [see *Powell*, 2007, and references therein], which allow the direct, model-free use of exogenous information in conditioning the operating policies [e.g., *Faber and Stedinger*, 2001; *Castelletti et al*., 2010, 2013; *Giuliani et al*., 2015]. The technical details about these methods are discussed in section 3.3. Notice that also the Basic Operating Policy can be designed by adopting the same approach by conditioning the decisions only on the time information.

In general, we expect the IOP to fill the performance gap between the upper and lower bound solutions (i.e., the POP and BOP) and, possibly, to produce a performance J^{IOP} as close as possible to J^{POP}. However, since the relative contribution of each component of $\mathbf{I}_t$ to the IOP performance might not be equivalent to the relative contribution in explaining the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$, we suggest to apply the ISA procedure in an iterative fashion (see Figure 1). At first, we consider only the candidate variable in Ξ_t with the highest ability in explaining the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$, assuming that it also has the highest potential to improve reservoir operations. We design an Improved Operating Policy conditioned on this variable only, and estimate the corresponding EVSI by comparison with the POP and BOP performance. We then iterate the procedure by incrementally adding variables to the surrogate information vector $\mathbf{I}_t$, designing the associated IOP, and evaluating the corresponding EVSI. When either the attained performance is satisfactory or the marginal improvement in the EVSI between two consecutive iterations is negligible, the procedure ends.

Again, in case of a single-objective problems, the EVSI is simply obtained as the difference between the (scalar) performances J^{POP} and J^{IOP} (possibly scaled by the basic performance J^{BOP}). In the multiobjective case, where the objective functions J^{POP} and J^{IOP} are vectors, the EVSI can be estimated by means of the quantitative metrics described in section 3.4.

3. Methods and Tools

In this section, we provide a short overview of the methods and tools that are available to perform the main steps of the ISA framework illustrated in the previous section.

3.1. Design of the Perfect Operating Policy

The Perfect Operating Policy can be reconstructed over a historical horizon by solving Problem (1) and assuming the sequence $\varepsilon_{[1,H]}$ is known. This is a standard nonlinear optimization problem and can be solved by either a local optimization method (e.g., gradient-based) or a global optimization method (e.g., direct search). Alternatively, if the objective function is time-separable, deterministic dynamic programming (DDP) can be used [*Bellman*, 1957]. Due to computational constraints, DDP can be applied only when the number of state and decision variables is sufficiently small (in the order of few units). When applicable, however, DDP provides an almost exact solution in a much more efficient fashion than other nonlinear optimization methods.

3.2. Automatic Selection of Information

The selection of the most valuable information to be used as a surrogate of the trajectory of future external drivers is reformulated here as the problem of finding the subset of variables $\mathbf{I}_t \subseteq \Xi_t$ that, together with t and $\mathbf{x}_t$, mostly explain the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$, which is obtained in the first step of our procedure. The rationale for including t and $\mathbf{x}_t$, which are not exogenous variables, is technical and aims to remove the part of the optimal sequence signal they explain. This facilitates the emergence of informative exogenous variables, which can be used in combination with t and $\mathbf{x}_t$ in the design of the Improved Operating Policies.

Especially when the number of candidate input variables is high, statistical techniques may help to perform the selection process in an automatic, reproducible way. While a standard cross-correlation analysis may fail in the presence of strongly nonlinear causal relationships, input variable selection (IVS) techniques can be used. IVS problems arise every time a variable of interest (i.e., the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$) has to be modeled as a function of a subset of potential explanatory variables (i.e., the exogenous variables $\mathbf{I}_t \subseteq \Xi_t$), but there is uncertainty about which subset to use among a number, usually large, of candidate variables available, which are often characterized by redundancy, collinearity, and highly nonlinear relationships [see *Galelli et al*., 2014, and references therein].

In the choice of the appropriate IVS method, three desirable features should be considered: (i) modeling flexibility, so to approximate strongly nonlinear functions, particularly because the functional relationship between the candidate inputs and the output $\mathbf{u}^{POP}_{[0,H-1]}$ is usually unknown a priori; (ii) computational efficiency, so to deal with potentially large data sets, when considering long-time series of observations and many candidate variables; and (iii) scalability with respect to the number of candidate input variables, so to handle numerous input variables with a different range of variability.

According to the guidelines provided in *Galelli et al*. [2014], in this paper, we use the hybrid model-based/model-free Iterative Input Selection (IIS) algorithm [*Galelli and Castelletti*, 2013a] combined with extremely randomized trees [*Geurts et al*., 2006; *Galelli and Castelletti*, 2013b], which possesses all the three properties discussed above. Given the output variable to be modeled $\mathbf{u}^{POP}_t$ and the set of candidate exogenous variables Ξ_t, the IIS algorithm first ranks these latter with respect to a statistical measure of significance and adds only the best performing input I_t^* to the set of selected variables $\mathbf{I}_t$. This operation aims to avoid the inclusion of redundant variables, as one an input is selected, all the inputs highly correlated with it may become useless in the next iterations. Then, the algorithm identifies a model of $\mathbf{u}_t^{POP}$ with input $\mathbf{I}_t$, namely $\mathbf{u}_t^{POP} = \hat{m}(\mathbf{I}_t)$, and computes the corresponding model performance with a suitable distance metric (e.g., the coefficient of determination) as well as the model residuals, which become the new output at the next iteration. The algorithm stops when the best variable returned by the rank is already in the set $\mathbf{I}_t$, or when overfitting conditions are reached. Further details on the IIS algorithm are provided in supporting information.

3.3. Design of the Improved Operating Policies

The use of traditional optimization techniques based on dynamic programming (DP) to design operating policies conditioned also on the selected information $\mathbf{I}_t$ might be limited by the so-called curse of modeling [*Tsitsiklis and Van Roy*, 1996]. DP indeed requires a model-based approach, where any information used to condition the operating policy must be explicitly modeled to fully predict the one-step ahead system transition used in the estimation of the value function. This information can be described either as a state variable of a dynamic model or as a stochastic disturbance, independent in time, with an associated pdf. As a consequence, exogenous information cannot be explicitly considered in conditioning the decisions, unless a dynamic model is identified for each additional variable, thus adding to the curse of dimensionality. The computational cost of DP indeed grows exponentially with the state vector dimensionality [*Bellman*, 1957] and, consequently, DP is not applicable when the dimensionality of the system exceeds two or three storages [*Loucks et al*., 2005; *Castelletti et al*., 2008b]. Moreover, the more complex the relationship between the exogenous variables (e.g., spatially correlated inflows), the lower will be the accuracy of the model and therefore its contribution toward improving the operating policy's performance.

To reduce the limiting effects of both the curse of modeling and the curse of dimensionality, we therefore need an approximate dynamic programming method that: (i) is scalable with respect to the state-decision space, thus overcoming the curse of dimensionality; (ii) is able to deal with a potentially large number of variables, possibly with different ranges of variability, and (iii) allows conditioning the operating policy on any exogenous information in a model-free fashion, namely without the need of explicitly modeling the temporal dynamics of these variables (thus overcoming the curse of modeling).

In the case study application, we adopt a policy search method [see *Deisenroth et al*., 2011, and references therein], as it represents a promising approach which possesses all the three properties discussed above. Policy search methods rely on a simulation-based optimization approach that first parameterizes the operating policy within a given family of functions and, then, optimizes the policy parameters (i.e., the decision variables of the problem) with respect to the operating objectives of the problem. This approach can be combined with any simulation model and allows the direct use of exogenous information available at time t to condition the decisions $\mathbf{u}_t$. In addition, this approach can be effectively combined with multiobjective evolutionary algorithms when the problem is characterized by high-dimensional decision spaces, noisy and multimodal objective functions, and multiple, competing operating objectives [*Giuliani et al*., 2015]. The mathematical formulation of the optimization problem solved by policy search is given in supporting information. In our case study application, the policy search approach was used also to design the Basic Operating Policy, by conditioning the decisions on the time information only.

3.4. Metrics to Assess the Value of Information

The quantification of the value of information for the analysis of both the EVPI and the EVSI is straightforward in single-objective problems, where it can be easily measured as the difference in the values of the (scalar) objective function considered. However, the majority of water resources management problems involves multiple competing objectives [e.g., *Kasprzyk et al*., 2009; *Giuliani et al*., 2014b]. Each operating policy is therefore associated to a vector $\mathbf{J} = [J^1, \ldots, J^K]$ of K different objectives, making the evaluation of the performance improvement more challenging. The solution of a multiobjective problem is indeed not unique, but rather a set of Pareto optimal (or approximate) solutions.

According to *Zitzler et al*. [2000], assessing the effectiveness of multiobjective problems' solutions requires to evaluate: (i) the distance of the final solutions from the true Pareto front; (ii) the coverage of the nondominated space; and (iii) the extent of the nondominated front. Among the commonly used metrics adopted in the literature [see *Maier et al*., 2014, and references therein], in this work we use the hypervolume

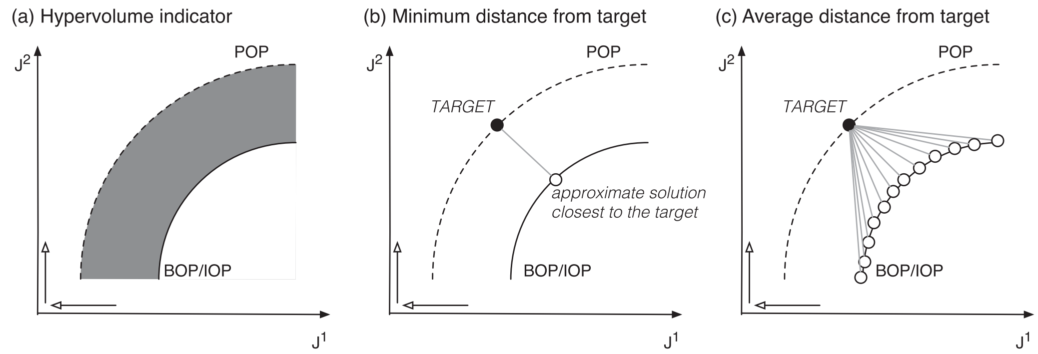

Figure 3. Illustration of the proposed metrics for assessing the value of information in multiobjective problems.

indicator *HV* as it captures both convergence and diversity [*Zitzler et al.*, 2003]. The hypervolume measures the volume of objective space dominated by an approximation set, with *HV* calculated as the difference in hypervolume between the best known Pareto optimal front (i.e., the set of POPs) and the considered approximation set (i.e., BOPs or IOPs), see Figure 3a. This metric allows for set-to-set evaluations, where the Pareto front with the higher *HV* is deemed the better.

The evaluation of the value of information should not be limited to the analysis of the whole Pareto front by means of *HV*. In fact, some information may be relevant for one (or few) objectives, whereas having no influence on other objectives. For example, precipitation measurements in the upstream catchment are potentially valuable for flood protection, while their role decreases for long-term objectives such as irrigation supply. As a consequence, adding information can alter the trade-off curve by producing a limited increase of *HV* but a significant improvement in a single objective, reducing the gap from some target optimal solutions.

To account for the specific advance toward a prespecified target solution, we introduce two additional metrics. The first metric measures the proximity between the target solution $\mathbf{J}^{POP,ref}$ and the closest point of the Pareto front under exam, see Figure 3b, i.e.,

$$D_{\min} = \min_{i=1,\ldots,N} ||\mathbf{J}^{POP,ref} - \mathbf{J}^i|| \tag{2}$$

where $||\cdot||$ stands for the (normalized) Euclidean norm, N is the number of points in the Pareto front under exam, and $\mathbf{J}^i$ is the objective vector representing the i-th point in such Pareto front. The lower D_{min}, the better the Pareto optimal set under exam. Since D_{min} is a point-to-point metric, achieving a good (small) value of D_{min} requires only a single solution in the Pareto optimal set close to the target solution.

The second metric, instead, is a set-to-point evaluation, which measures the average distance of the entire Pareto front under exam from the target POP, see Figure 3c, i.e.,

$$D_{avg} = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{J}^{POP,ref} - \mathbf{J}^i|| \tag{3}$$

Again, the lower the value of D_{avg}, the better the set of Pareto optimal solutions under exam. The underlying idea is that the potential changes in the trade-offs curve produced by the additional information may also modify the structure of preferences of the decision maker. In this case, he/she might be interested in obtaining not only a single solution very close to the target POP but also a set of solutions around the target so to explore the new trade-offs generated by using the additional information.

4. The Hoa Binh Reservoir Case Study

The Hoa Binh reservoir system (Figure 4) is a multipurpose regulated reservoir in the Red River basin (Vietnam). The Red River is the second largest basin of Vietnam and is shared by China (48%), Vietnam (51%), and the rest in Laos. It has three major tributaries, namely the Da River, the Thao River, and the Lo River. The Da River is the most important water source, contributing for 42% of the total discharge that enters in the downstream part of the basin, ultimately flowing through the capital city of Hanoi. The Hoa Binh

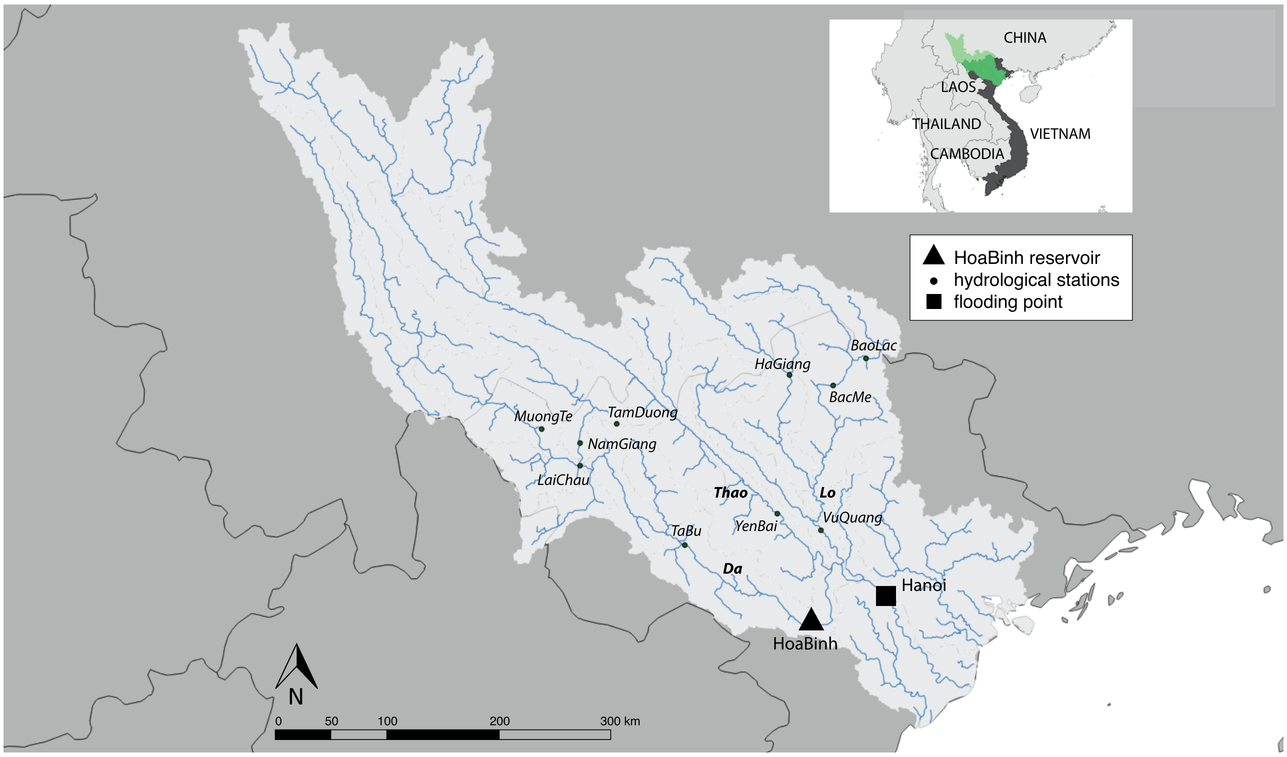

Figure 4. Map of the Red River basin.

reservoir, constructed on the Da River, has a surface area of about 198 km^2 and an active storage capacity of about 6 billion m^3. Given this large storage capacity, the Hoa Binh regulation plays a key role for flood mitigation in the downstream part of the Red River catchment and, especially, in the Hanoi metropolitan area, where 6.5 million of people live. The Hoa Binh reservoir is also connected to a power plant equipped with eight turbines, for a total design capacity of 1920 MW, which corresponds to the 19% of the installed national hydropower capacity and contributes a large share of the national electricity production.

4.1. Model

The system is modeled by a combination of conceptual and data-driven models with a daily time resolution. The dynamics of the Hoa Binh reservoir is described by the mass balance equation of the water volume s_t stored in the reservoir, i.e.

$$s_{t+1} = s_t + q_{t+1}^{Da} - r_{t+1} \tag{4}$$

where s_t is the reservoir storage, q_{t+1}^{Da} is the net inflow in the interval $[t, t+1)$ (obtained by mass balance inversion), and r_{t+1} is the volume released in the same interval. The release is defined as $r_{t+1} = f(s_t, u_t, q_{t+1}^{Da})$, where $f(\cdot)$ describes the nonlinear, stochastic relation between the decision u_t, and the actual release r_{t+1} [*Piccardi and Soncini-Sessa*, 1991]. The release r_{t+1} coincides with the release decision u_t unless a correction is needed in order to take into account the legal and physical constraints on the reservoir level and release, including spills when the reservoir level exceeds the maximum capacity.

The electricity production of the hydropower plant is determined by a function of the simulated reservoir release r_{t+1} and the net hydraulic head (i.e., reservoir level minus tailwater level). The water level in Hanoi is estimated by a flow routing model consisting of a data-driven feedforward neural network fed by the Hoa Binh release (r_{t+1}), the discharge from the Thao River, measured at Yen Bai (q_{t+1}^{YB}), and the discharge from the Lo River, measured at Vu Quang (q_{t+1}^{VQ}). Further details about the model of the Hoa Binh system and its operations can be found in *Castelletti et al.* [2012]; *Pianosi et al.* [2012]; and *Castelletti et al.* [2013].

The two main objectives of the Hoa Binh regulation are hydropower production and flood control. They are formalized as follows:

1. *Hydropower production* (J^{hyd}): the daily average hydropower production (kWh/d) at the Hoa Binh hydropower plant, to be maximized, defined as

$$J^{hyd} = \frac{1}{H}\sum_{t=0}^{H-1} HP_{t+1} \quad \text{with } HP_{t+1} = \left(\eta g \gamma_w \bar{h}_t q_{t+1}^{Turb}\right) \cdot 10^{-6} \tag{5}$$

where η is the turbine efficiency, $g = 9.81$ (m/s^2) the gravitational acceleration, $\gamma_w = 1000$ (kg/m^3) the water density, $\bar{h}_t$ (m) the net hydraulic head (i.e., reservoir level minus tailwater level), and q_{t+1}^{Turb} (m^3/s) the turbined flow;

2. *Flood damages* (J^{flo}): the daily average excess level h_{t+1}^{Hanoi}(cm^2/d) in Hanoi with respect to the flooding threshold $\bar{h} = 950$ cm, to be minimized, defined as

$$J^{flo} = \frac{1}{H}\sum_{t=0}^{H-1} max\left(h_{t+1}^{Hanoi} - \bar{h}, 0\right)^2 \tag{6}$$

where h_{t+1}^{Hanoi} is the level in Hanoi estimated by the ANN flow routing model, which depends on the Hoa Binh release (r_{t+1}) along with the Thao (q_{t+1}^{YB}) and Lo (q_{t+1}^{VQ}) discharges.

In summary, our model is composed by two competing objectives, one state variable representing the Hoa Binh storage s_t, one release decision u_t, and a vector of uncontrolled external drivers comprising the three inflows $\mathbf{q}_{t+1} = [q_{t+1}^{Da}, q_{t+1}^{YB}, q_{t+1}^{VQ}]$. The same model is used in all the steps of the framework for contrasting the performance of the Perfect, Basic, and Improved Operating Policies. Although this system represents a relatively simple problem, the presence of multiple inflows combined the lack of reliable forecasts due to the missing information on the upstream Chinese part of the catchment makes the case study particularly suitable for demonstrating the potential of our ISA framework, while being sufficiently well understood to also allow for a robust interpretation of the results. In addition, two of the three inflows are located downstream with respect to the Hoa Binh reservoir and do not impact on the reservoir dynamics and on the hydropower production. Yet, they contribute more than 50% of the flow in the downstream part of the system, thus reducing the buffer potential of the Hoa Binh reservoir during the monsoon season, ultimately increasing the risk of flood in Hanoi.

4.2. Experiment Setting

1. *Observational data*: time series of rainfall, temperature, and streamflow at various locations in the catchment and rivers network are available at daily resolution starting from the 1950s. Starting from the late 1980s and following the construction of the Hoa Binh reservoir, which was completed in 1994, time series of reservoir levels, releases, and net inflows are also available. Since the full reservoir operations started only in 1995 (when it was completely filled up), we consider the time horizon 1995–2006 as in *Castelletti et al*. [2012] to demonstrate and validate our framework. This time horizon is used for both the optimization and evaluation of the policy performance, thus ensuring that the Perfect Operating Policies represent the upper bound in terms of system performance. In particular, the approximate the values of J^{hyd} and J^{flo}, which would require to simulate the system under an infinite number of disturbance realizations, each of infinite length, by computing the sample average value using the 12 years time series of the historical disturbances [see *Pianosi et al*., 2011 and discussion in supporting information].
2. *Perfect Operating Policies*: the set of POPs was designed via Deterministic Dynamic Programming, with an 11 years management horizon from 1995 to 2006. The weighting method [*Gass and Saaty*, 1955] is used to convert the 2-objective problem into a single-objective one via convex combination. Eleven combinations of weights were evaluated in this study as they provide a good exploration of the trade-offs curve.
3. *Improved Operating Policies*: the set of IOPs was designed via evolutionary multiobjective direct policy search (EMODPS), a simulation-based optimization approach which overcomes the limitations of DP family methods by combining direct policy search, nonlinear approximating networks, and multiobjective evolutionary algorithms [*Giuliani et al*., 2015]. In particular, it allows the direct use of exogenous information in conditioning the decisions as well as the estimation of an approximation of the Pareto front in a

single run of the optimization algorithm. The drawback of this method is that it is a heuristic approach and there is no theoretical guarantee on the optimality of the resulting solutions. Their accuracy strongly depends on the choices of the class of functions used to parameterize the operating policy and on the efficiency of the algorithm used to optimize the policy parameters. In this work, we use Gaussian Radial Basis Functions (RBFs) to parameterize the operating policy as they are capable of representing functions for a large class of problems [e.g., *Mhaskar and Micchelli*, 1992; *Busoniu et al*., 2011] and have been demonstrated to be more effective than other universal approximators [*Giuliani et al*., 2014c,2015]. The number of basis of the RBFs is set equal to $N=M+1$, where M is the dimension of the policy input vector $\mathcal{I}_t=(t, \mathbf{x}_t, \mathbf{I}_t)$. The corresponding number of policy parameters to be optimized is $n_\theta=N(2M+1)$. To perform the optimization, we use the self-adaptive Borg MOEA [*Hadka and Reed*, 2013], which has been shown to be highly robust across a diverse suite of challenging multiobjective problems, where it met or exceeded the performance of other state-of-the-art MOEAs [*Hadka and Reed*, 2012; *Reed et al*., 2013]. Since the Borg MOEA has been demonstrated to be relatively insensitive to the tuning parameters, we use the default algorithm parameterization suggested by *Hadka and Reed* [2013], overcoming the limitations of tuning the algorithm parameters to the specific problem. Each optimization was run for 250,000 function evaluations, with the simulation of the system performed over the same horizon of DDP (i.e., 1995–2006). To improve solution diversity and avoid dependence on randomness, the solution set from each formulation is the result of 20 random optimization trials. The final set of Pareto optimal policies for each experiment is defined as the set of nondominated solutions from the results of all the optimization trials.

4. *Basic Operating Policies*: the set of BOPs is defined as a set of open-loop policies conditioned only on the day of the year (i.e., release plans), designed via EMODPS as the Improved Operating Policies. We do not consider the historical policy of the Hoa Binh as a benchmark for the following reasons: our model considers only two objectives, while in reality the Hoa Binh reservoir has been regulated also for irrigation supply and for ensuring navigation in the Red River delta; the historical regulation before 1995 was also conditioned by the fact that reservoir filling was not complete, while after 2005 it was also affected by the undergoing construction of Son La reservoir upstream of the Hoa Binh.
5. *Automatic selection of information*: the identification of the most relevant information was carried out by using the hybrid model-based/model-free Iterative Input Selection (IIS) algorithm [*Galelli and Castelletti*, 2013a]. The IIS was run on a sample data set comprising the day of the year (t), the Hoa Binh storage (s_t), and a set Ξ_t including 12 exogenous variables: the Hoa Binh net inflow (q_t^{Da}), the spatial average precipitation in the Da River catchment ($\tilde{P}_t^{Da}$), flows (q_t^i) and precipitations (P_t^j) measured at different locations ($i=$ Ta Bu, Lai Chau, Nam Giang, Yen Bai, Vu Quang and $j=$ Muong Te, Tam Duong, Bao Lac, Bac Me, Ha Giang, see Figure 4).

5. Results

5.1. Quantifying the EVPI

The first step of the ISA framework (Figure 1) is the estimation of the Expected Value of Perfect Information by contrasting the Perfect Operating Policies (POPs) and the Basic Operating Policies (BOPs). Figure 5a shows the performance of the POPs (represented by black squares) evaluated over the horizon 1995–2006. In the same figure, the red points represent the performance of the BOPs, while the performance of the Improved Operating Policies (i.e., blue, green, magenta points) will be discussed later on in section 5.2. The arrows indicate the direction of increasing preference, with the best solution located in the top-left corner of the figure. Visual comparison of the Pareto fronts shows that the potential space for improvement generated by the knowledge of perfect information of the future inflows trajectories, which is available in the problem formulation of the POPs and not in the BOPs' one, is relevant in terms of both the operating objectives (see the area between the black squares and the red points).

According to the shape of the resulting Pareto front, we focus the analysis on a single target Perfect Operating Policy, which represents a potentially good compromise between J^{hyd} and J^{flo}. Beyond representing a good target solution, previous works [*Castelletti et al*., 2012] have shown that this compromise solution is particularly difficult to be translated into an actual operating policy. In fact, while the two extreme policies that would optimize the two single objectives separately are relatively simple to design (hydropower

10.1002/2015WR017044

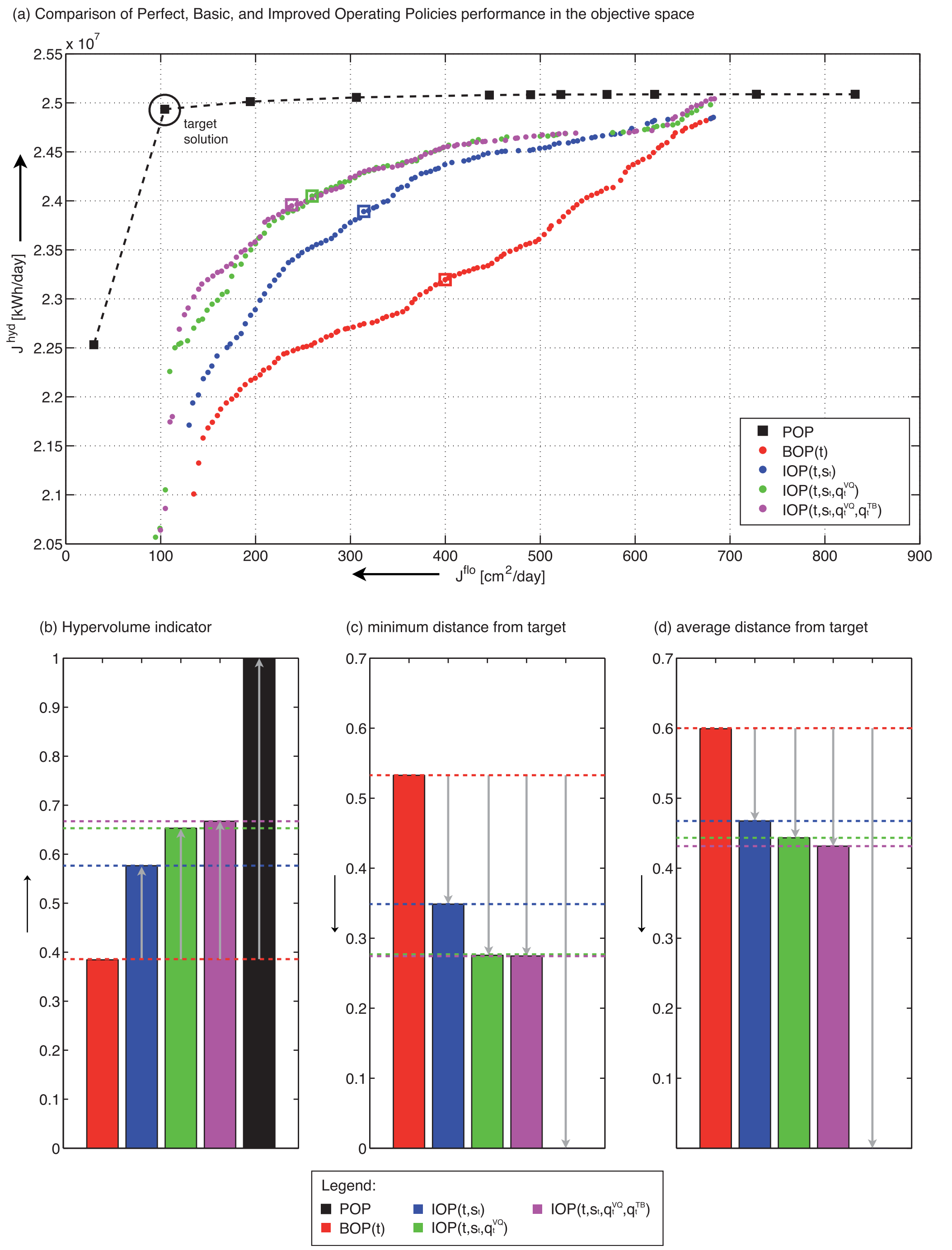

Figure 5. Comparison of the performance obtained by Perfect Operating Policies, Basic Operating Policies, and three Improved Operating Policies conditioned on increasing information (i.e., time and storage in blue; time, storage, streamflow at Vu Quang in green; time, storage, streamflow at Vu Quang, streamflow at Ta Bu in magenta). (a) Shows the policies' performance in the objective space, (b–d) report the value of information quantified by the three metrics introduced in section 3.4.

production is easily maximized by maintaining the release equal to the turbine capacity, whereas flood protection is maximized by maintaining the reservoir level as low as possible outside flood events), designing the compromise policy that would ensure adequate flood protection while simultaneously maintaining high hydropower production is particularly challenging. This observation is confirmed in Figure 5a, where the two extremes of the red Pareto front are not far from the corresponding solutions in the POPs' set, while the maximum distance between the two fronts is located in correspondence with

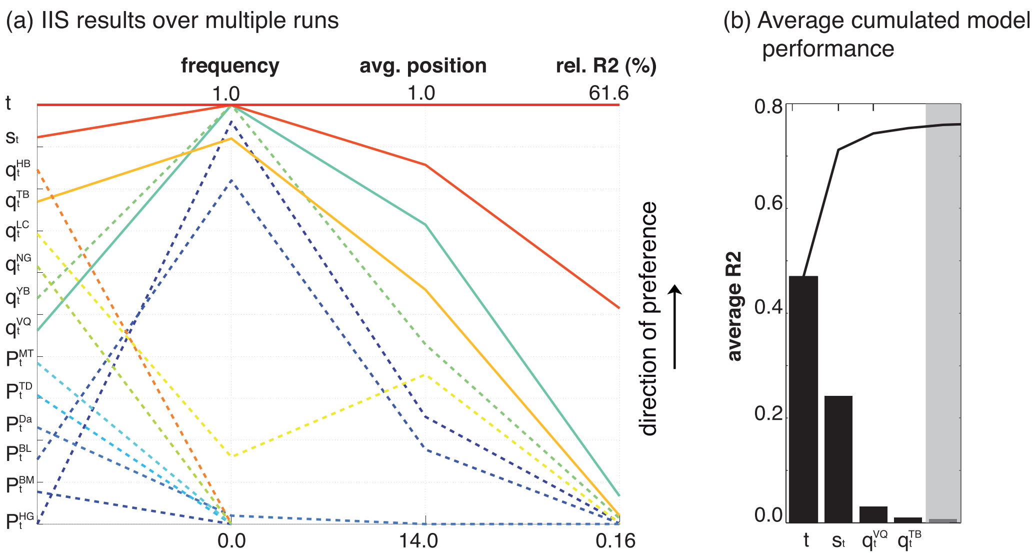

Figure 6. Information selection results obtained via 50 runs of the IIS algorithm: (a)for each candidate variable, the frequency of selection, the average position, and the average relative contribution (t is day of the year; s_t is the Hoa Binh storage, q^i and P^i are measured flow and precipitation at different stations reported in Figure 4); (b) the average cumulated performance of the regression model describing the optimal decisions sequence $u^{POP}_{[0,H-1]}$.

the selected target solution. This motivates searching for a set of Improved Operating Policies (IOPs) that can fill in such space.

A more quantitative assessment of the EVPI is provided by the values of the three metrics introduced in section 3.4, where the POPs represent the optimal front and $\mathbf{J}^{POP,ref}$ is the target POP. Figure 5b shows that the difference in the hypervolume indicator between the Basic Operating Policies and the POPs is 0.62, which confirms the large gap between the set of perfect and basic operating policies. In addition, the BOPs fail in exploring the trade-offs region around the target solution, attaining large values in both D_{min} and D_{avg} (Figures 5c and 5d).

5.2. Information Selection

The second step of the ISA framework (Figure 1) aims at identifying the best subset of information $\mathbf{I}_t \subseteq \Xi_t$ that can be used together with the day of the year t and the state vector $\mathbf{x}_t$ in the design of IOPs. As anticipated, we used the IIS algorithm to select the most relevant information to explain the optimal sequence of release decisions $u^{POP}_{[0,H-1]}$ from a sample data set of 12 candidate exogenous variables, along with the day of the year (t) and the Hoa Binh storage (s_t). The rationale for including these latter is to avoid the selection of exogenous variables correlated with t or s_t, which might be ineffective as surrogates of the future external drivers.

Figure 6a reports the results of 50 runs of the IIS algorithm using a parallel axes plot. The repetition of the experiments aims at filtering the randomness associated to the construction of the extra-trees models used by the IIS algorithm [*Galelli and Castelletti*, 2013a]. In this representation, each variable is represented as a line (also identified by the different colors) crossing the three axes at the values of the corresponding performance in terms of frequency of selection, average position, and average relative contribution over the 50 runs. In the plot, the reported performance is normalized between their minimum and maximum values and the axes are oriented so that the direction of preference is always upward. Consequently, the most relevant variables would be represented by horizontal lines running along the tops of all of the axes, meaning variables that are selected with high frequency, in the first positions, and with the largest relative contributions.

The results in Figure 6a show limited variability over the 50 runs: the three variables providing the highest relative contributions (i.e., t, s_t, and q_t^{VQ}) are selected in all the 50 runs in the same positions. Figure 6b shows the average performance attained by the regression model in describing the optimal release decisions sequence $u^{POP}_{[0,H-1]}$, measured in terms of cumulated coefficient of variation (R2). From these results, we tentatively stop the selection of the relevant information at the first four selected variables because they allow explaining a sufficiently high percentage of the output variance (i.e., R2 = 75%). The four variables selected, namely day of the year (t), Hoa Binh storage (s_t), streamflow at Vu Quang (q_t^{VQ}), and streamflow at Ta Bu (q_t^{TB}), are the ones represented by solid lines in Figure 6a. Another potentially relevant variables might be the flow at Yen Bai (q_t^{YB}), which is selected in all the runs in the fifth position (while q_t^{TB} is selected in fourth position in 46 over 50) and provides the fifth highest contribution. However, the high correlation between the flow in the Thao River with the ones in the Da (86.9%) and Lo (79.8%), and the minor contribution of the Thao River to the flow in Hanoi (i.e., around 19%, while the Da and Lo rivers account for 42% and 25.4%, respectively) are likely reducing the value of q_t^{YB}. The need to consider also q_t^{YB} to better condition the Hoa Binh operations will be determined on the basis of the resulting performance of the operating policies conditioned on the four selected inputs. Finally, it is worth noting that when q_t^{TB} is not selected, it is actually replaced by the flow at Lai Chau (q_t^{LC}). This latter is still a measure of the flow in Da River, observed upstream with respect to Ta Bu station (see Figure 4), and it is likely highly correlated with q_t^{TB}.

The selection of this set of variables is interesting at least with two respects: (i) the Hoa Binh inflow (q_t^{HB}) is never selected, probably because the optimal operations of the reservoir requires some forms of anticipation to effectively face the inflow dynamics, which is better captured by observations in the upstream part of the Da River catchment (e.g., flow at Ta Bu); (ii) the flow in the Lo River at Vu Quang (q_t^{VQ}) is selected before the one in the Da River at Ta Bu (q_t^{TB}), meaning that the information on the uncontrolled part of the system is more relevant for flood protection in Hanoi than a refined predictions of the reservoir inflow.

5.3. Improved Operating Policies and Assessment of the EVSI

In this section, we incrementally add the variables selected by the IIS algorithm to the set of policy input and we iteratively design and analyze the resulting Improved Operating Policies both in terms of their performance with respect to the two operating objectives (i.e., J^{hyd} and J^{flo}) and to the associated Expected Value of Sample Information (step 3 of the ISA framework, see Figure 1).

The first variable selected is the day of the year (t). This is not surprising given the strong influence of the seasonality and of the monsoon period on the system dynamics. The set of policies conditioned on the day of the year is the already discussed set of Basic Operating Policies, represented by the red points in Figure 5a. The second selected variable is the Hoa Binh storage (s_t). The associated Improved Operating Policies IOP(t,s_t) are represented by blue points. It is worth noting that the set (t,s_t) is the minimum information generally used for the design of closed loop operating policies [*Bertsekas*, 1976]. The comparison of the performance of BOP(t) and IOP(t,s_t) shows a large contribution associated to the Hoa Binh storage, demonstrating the advantage of closing the loop between operational decisions and evolving system conditions. A quantitative evaluation of the EVSI is given by the metrics reported in Figure 5b: *HV* increases from 0.38 to 0.57 (i.e., +50%) when moving from BOP(t) to IOP(t,s_t), while D_{min} and D_{avg} decrease from 0.53 to 0.34 (i.e., −35%) and from 0.59 to 0.46 (i.e., −22%), respectively. While the large increase of *HV* suggests a general improvement of the Pareto front enhancing both the objectives, the changes in the other two metrics are relatively smaller. This means that conditioning the Hoa Binh policy on the reservoir storage is not sufficient to approach the specific target POP solution (Figure 5a, black square). These results motivate for further improving the operating policy by introducing additional information to better anticipate the monsoon season and reduce flooding in Hanoi.

The third variable selected is the previous day flow in the Lo River measured at Vu Quang (q_t^{VQ}). The resulting performance of the new set of policies IOP(t, s_t, q_t^{VQ}) is represented in Figure 5a by green points. Results show that the marginal improvement obtained by adding q_t^{VQ} is lower than the one obtained by including the Hoa Binh storage (s_t). However, although the maximum of hydropower production is almost the same, the green solutions make a relevant step toward the target solution, with significant improvements in the left part of the Pareto front. The values of the metrics (Figures 5b–5d) confirm this visual evaluation and the point-to-point metric D_{min} is the one with the largest improvement (larger than 20%, while *HV* and D_{avg} improvements are equal to 14% and 2%, respectively).

Finally, the last variable selected is the previous day flow in the Da River measured at Ta Bu (q_t^{TB}). The performance of the IOPs($t, s_t, q_t^{VQ}, q_t^{TB}$) is represented in Figure 5a by magenta points. Results show that the EVSI of the flow observed at Ta Bu is marginal and only allows the attainment of higher hydropower production for values of J^{flo} between 100 and 200 cm^2/d. The values of the metrics confirm this qualitative observation, attaining an average improvement among the three metrics smaller than 2%.

Given the limited improvement obtained by adding q_t^{TB}, we decided to stop the policy design at this step. A large gap still remains between the best Improved Operating Policies IOP($t, s_t, q_t^{VQ}, q_t^{TB}$) and the set of Perfect Operating Policies. However, such a gap could only be closed by using information able to "anticipate" floods by at least 4 days, such as observations of precipitation or streamflow in the upstream Chinese part of the Da River basin or accurate streamflow forecasts. In fact, previous studies [*Castelletti et al.*, 2012] show that this is the time needed to drawdown the Hoa Binh level and to create the storage volume for large flood events. Such information is currently not available due to the lack of hydrometeorological data in the upstream Chinese part of the Da River basin.

The values of the metrics used for assessing the EVSI can also be compared to the results of the IIS algorithm to understand whether there exists a direct relationship between the performance of the regression model in explaining the optimal sequence of release decisions and the performance of the associated operating policies. Results show that increasing information yields higher performance both in terms of R^2 and EVSI. We observe the presence of a saturation effect, which confirms that the day of the year (t) and the Hoa Binh storage (s_t) represent the most valuable information (i.e., $R^2=0.71$, $D_{min}=0.35$, and $HV=0.58$) while the contribution of the other two variables, namely the previous day streamflow at Vu Quang (q_t^{VQ}) and at Ta Bu (q_t^{TB}), is smaller (i.e., $\Delta R^2=0.04$, $\Delta D_{\mathrm{min}}=0.07$, and $\Delta HV=0.09$). However, the relationships between the regression model's performance and the associated values of the three metrics assessing the value of information seem to be nonlinear, thus suggesting that a given improvement in explaining the optimal sequence of release decisions does not yield the same improvement in the performance of the set of operating policies.

5.4. Analysis of the Operating Policies

To better understand the contribution of each selected variable in enhancing the operations of the Hoa Binh reservoir, we analyze the dynamic behavior of the system under different operating policies conditioned over distinct information. The selected solutions, identified by the boxed points in Figure 5a, are: the target Perfect Operating Policy (black), one Basic Operating Policy selected from the set of BOPs as being closest to the target solution (red), and one Improved Operating Policy selected from each respective set according to the same criterion (blue, green, magenta).

Figure 7 reports the simulated trajectory of the Hoa Binh storage under each considered policy. As clearly shown by the cyclostationary mean over the period 1995–2006 represented in Figure 7a, the timing of the reservoir drawdown is key to effectively balance J^{hyd} and J^{flo}. The target Perfect Operating Policy (black line) is indeed able to delay the reservoir drawdown as much as possible and to refill it as soon as possible in order to reduce the losses in hydropower generation that are encountered when the reservoir level must be kept low during the flood season. The Basic Operating Policy (red line) shows two large drawdown cycles, corresponding to the two peaks of the monsoon, with the first one operated more than 1 month early than under the target solution. This is because the BOP does not use any variable describing the current system condition and therefore is overly conservative. The trajectories obtained under the considered Improved Operating Policies become closer and closer to the ideal black one when moving from the simplest alternative (blue) to the most sophisticate one (magenta).

Figures 7b–7e show the storage trajectories in the individual years (the color tone is used to differentiate the years from 1995 (dark) to 2006 (light)). The main differences from one policy to another can be observed in the two drawdown cycles. Both the Basic and the Improved Operating Policies (Figures 7c–7f) show similar patterns during the first drawdown (i.e., from day 100 to 150) in all years. On the contrary, the target Perfect Operating Policy (Figure 7b) generates different trajectories, and in particular different timing of the drawdown, across the years: for instance, for 2 years when no large floods occurred the reservoir is maintained full until day 130. A similar behavior is observed in the second cycle (from day 200 to 250). The Basic Operating Policy produces significant drawdowns every year (Figure 7c), whereas the trajectories tend to differentiate across the years when additional information is considered (Figures 7d and 7e). In particular, the IOP($t, s_t, q_t^{VQ}, q_t^{TB}$) reported in Figure 7e effectively reproduces the trajectories of the target POP, with larger drawdown in the first (dark) years than in the last (light) ones.

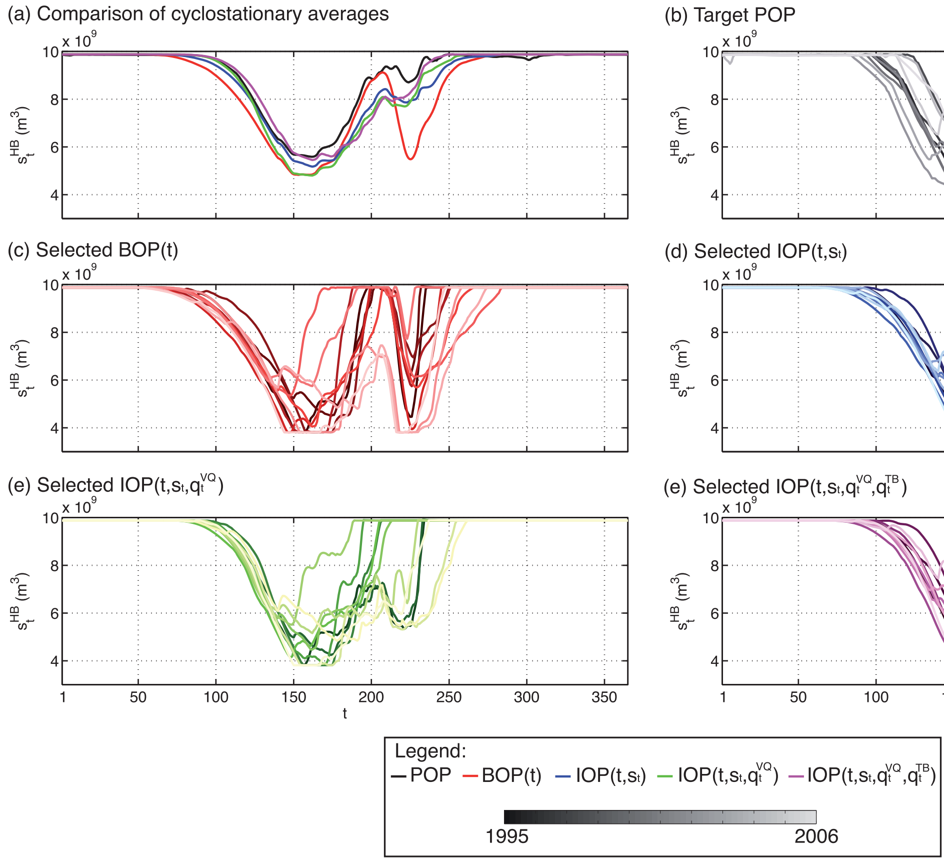

Figure 7. Comparison of the Hoa Binh storage trajectories under the target Perfect Operating Policy, the selected Basic Operating Policy, and the selected Improved Operating Policies.

Finally, Figure 8 compares the simulated level in Hanoi during a flood event in 2002. The black line shows the ideal trajectory obtained under the target Perfect Operating Policy. According to the objective formulation which elevates the level excess to the square (see equation (6)), this solution successfully minimizes J^{flo} by reducing the flood peak and distributing the flood over time. All the other policies, which do not have perfect information on the future, are instead less effective and produce higher water levels, particularly

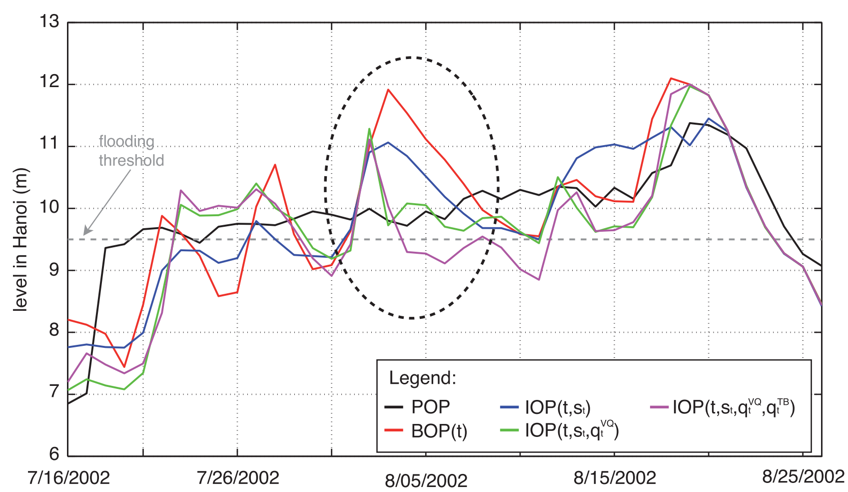

Figure 8. Comparison of the Hanoi level during a flood event in 2002 under the target Perfect Operating Policy, the selected Basic Operating Policy, and the selected Improved Operating Policies.

Table 1. Economic Value of Information

Selected Policy	J^{hyd} (kWh/d)	ΔJ^{hyd} (kWh/d)	J^{flo} (cm^2/d)	ΔJ^{flo} (cm^2/d)	Economic Flood cost (US$/d)	Economic Value of Information (US$/d)
BOP(t)	2.32×10^7		399.8		5117.6	
IOP(t,s_t)	2.38×10^7	+2.6%	314.0	−21.5%	4019.4	1098.2
IOP(t, s_t, q_t^{VQ})	2.40×10^7	+3.4%	259.6	−35.1%	3322.6	1795.0
IOP($t, s_t, q_t^{VQ}, q_t^{TB}$)	2.39×10^7	+3.0%	238.0	−40.5%	3046.3	2071.3
POP	2.49×10^7	+7.3%	104.3	−73.9%	1334.6	3783.0

during the two peaks on 5 August and 20 August. The trajectories during the first peak further demonstrate the contribution of each selected variable in approximating the target POP solution (see the dashed black circle). The red trajectory obtained under the Basic Operating Policy produces the highest water levels, which explains its poor performance in terms of J^{flo}. As in Figure 7a, the Improved Operating Policies become closer and closer to the target POP by incrementally adding information to condition the operating policy.

5.5. The Economic Value of Information

The earlier analysis can be integrated to provide an estimate of the economic value of the information used in conditioning the Hoa Binh operations, which can be defined by the reduction in the flood costs attained by using a larger set of information. This monetary value represents an indication about the maximum amount a decision maker might be willing to pay in order to acquire the information needed for conditioning the policy and increase the flood protection. We focus the analysis on the same four policies considered in the previous section, namely the target Perfect Operating Policy along with one Basic and one Improved Operating Policy selected from each respective set as being closest to the target solution (see Figure 5a, the boxed points).

Since the two considered operating objectives are not expressed in monetary values, the economic cost of flooding is estimated by adopting the same approach used in *Giuliani and Castelletti* [2013], which relies on the a posteriori trade-off analysis of the Pareto front and the evaluation of the corresponding shadow prices, which provide an estimate of the flood damages' marginal abatement costs [*Lee et al*., 2002]. In the Hoa Binh problem, the definition of Pareto optimality ensures that the decrease in the hydropower production associated to a given Pareto optimal solution with respect to the hydropower extreme of the Pareto front (i.e., the solution designed assigning weight equal 1 to J^{hyd} and equal 0 to J^{flo}) is compensated by the corresponding reduction in the flooding objective. For example, the difference in terms of hydropower production attained by the target Perfect Operating Policy with respect to the hydropower extreme of the POPs' Pareto front is $\Delta J^{hyd} = -153{,}171$ kWh/d. This worsening is however balanced by a better performance in terms of J^{flo}, namely $\Delta J^{flo} = -727.5$ cm^2/d.

By relying on this idea, the economic cost of flooding can be estimated by replacing the hydropower production with the associated revenue. In particular, we assume an average energy price equal to 0.06 US$/kWh, which does not change during the year. Note that this is not the real energy price in Vietnam because there is no energy market. This value has been estimated on the basis of the price applied to the energy imported in Vietnam from China. The flooding cost v can be estimated as the slope of the line connecting the target solution to the hydropower extreme of the Pareto front, i.e.

$$v = \frac{\Delta J^{hyd} \times 0.06}{\Delta J^{flo}} = 12.80 \quad \frac{US\$}{cm^2} = 345.27 \quad \frac{US\$}{cm} \tag{7}$$

The estimated flood cost v allows an economic assessment of the flood costs' reductions potentially achievable under the different operating policies. Table 1 reports the performance attained by each policy (i.e., BOP, IOPs, and POP) in the two operating objectives, the associated economic cost of flooding, and the estimated economic value of information. The potential improvements achievable with the Perfect Operating Policy with respect to the Basic Operating Policy (i.e., 7.3% in terms J^{hyd} and 73.9% in terms of J^{flo}) confirm the challenges associated to the selected target solution as well as the focus of the overall analysis on guaranteeing adequate flood protection. The results attained by the Improved Operating Policies demonstrate the effectiveness of the proposed procedure, which provides an improvement of 40% in terms of J^{flo} with the IOP($t, s_t, q_t^{VQ}, q_t^{TB}$).

In addition, the computation of the economic value of information allows associating monetary values to the information employed in conditioning the Hoa Binh operating policy. Numerical results confirm that a major contribution is provided by the use of the Hoa Binh storage, which is associated to a 20% reduction of the flood cost and an economic value of information equal to 1098.2 US$/d. The economic value of information then increases when additional variables are considered, even though the marginal improvement tend to decrease when moving from the IOP(t,s_t) to the IOP($t, s_t, q_t^{VQ}, q_t^{TB}$). Finally, a large gap remains between this latter and the Perfect Operating Policy (i.e., around 1700 US$/d), which further confirms that the optimal operations of the Hoa Binh would require additional information that is not currently available, such as observations of hydrometeorological data in the upstream Chinese part of the Da River basin.

6. Conclusions

In this paper, we propose an Information Selection and Assessment (ISA) framework to automatically select the most valuable information for informing water systems operations. We suggest several computational tools that can be used to directly embed this information into the design of reservoir operating policies and we provide quantitative metrics to assess the operational and economic value of information. The regulation of the Hoa Binh water reservoir in Vietnam is used as a case study.

Results show the effectiveness of the proposed ISA framework: by incrementally adding variables to the reservoir operating policies, the corresponding Pareto fronts move toward the set of Perfect Operating Policies (POPs) obtained under the assumption of perfect information on the future. Since in the proposed application a single target POP solution is used to select the information, we observe larger improvements close to this target solution than in other part of the Pareto front. The best Improved Operating Policies indeed attain a 3% increase in hydropower production and a 40% improvement in flood protection. This seems to be reasonable as the designed operating policies are conditioned upon information that is relevant for one specific compromise between hydropower production and flood control, while different variables might be more significant for other target solutions.

Further research efforts will focus on the application of the ISA framework to reservoir operations problems involving long-term objectives, for instance irrigation supply, along with the use of more complex data sources, such as snow information, remote sensed data, or weather forecasts. While the steps and methods employed in the ISA would remain exactly the same, we expect its value should increase given the traditional difficulties of identifying effective exogenous variables that could convey information relevant over a long lead time. This should be possible for instance in those regions where low-frequency climate phenomena, such as El Nino Southern Oscillation, seem to have a significant impact on local climate and hydrology [e.g., *Hamlet and Lettenmaier*, 2000; *Beltrame et al*., 2014].

Another next step of our research is the comparison of the model-free ISA framework and the more conventional approach where observations are used to force a flow forecasting model, and the flow forecasts are then used to inform reservoir decisions. This approach might provide a better anticipation capacity and capture information about evolving conditions (e.g., change in flow velocity linked to flood wave, optimization of choice of optimal time of observation in order to capture specific part of the rising limb, peak, and recession phase), which are not taken into consideration in the model-free ISA framework. This comparison goes beyond the scope of the present paper, which focuses on introducing and discussing the former approach. A consistent comparison of the two approaches across catchments with different characteristics (e.g., snowmelt-dominated, arid or semiarid, etc.) and management problems with different objectives (e.g., long-term or short-term) could help understanding the links between such features of a water system and the likely benefit of developing a flow forecasting system versus the direct use of information in water systems operations. These insights could help in setting priorities between the improvement of monitoring systems or, on the contrary, of flow forecasting systems. It is worth noting that the ISA framework can also be potentially used to support the identification of the most valuable forecasts (e.g., the best/maximum lead time), particularly in the case of medium-long-term forecasts.

Finally, the ISA framework can be used for improving the operations of any water system, from water distribution to groundwater pumping, and can be further extended in order to assess the value of information with respect to multiple competing objectives and to study how changing the trade-off (i.e., the target solution) requires using different information.

Acknowledgments
This work was partially supported by the *IMRR - Integrated and sustainable water Management of the Red-Thai Binh Rivers System in changing climate* research project funded by the Italian Ministry of Foreign Affair as part of its development cooperation program. Francesca Pianosi was supported by the Natural Environment Research Council (Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1). The authors would like to thank Alessia Cavalli and Lea Janossy for their contribution in developing initial numerical experiments.The data used in the study are from the Ministry of Agriculture and Rural Development (MARD) of Vietnam and have been collected during the IMRR project (http://http://xake.elet.polimi.it/imrr).

References

Alberini, A., A. Hunt, and A. Markandya (2006), Willingness to pay to reduce mortality risks: Evidence from a three-country contingent valuation study, *Environ. Resour. Econ.*, *33*(2), 251–264.

Alemu, E., R. Palmer, A. Polebitski, and B. Meaker (2011), Decision support system for optimizing reservoir operations using ensemble streamflow predictions, *J. Water Resour. Plann. Manage.*, *137*(1), 72–82, doi:10.1061/(ASCE)WR.1943-5452.0000088.

Anghileri, D., A. Castelletti, F. Pianosi, R. Soncini-Sessa, and E. Weber (2013), Optimizing watershed management by coordinated operation of storing facilities, *J. Water Resour. Plann. Manage.*, *139*(5), 492–500.

Ansar, A., B. Flyvbjerg, A. Budzier, and D. Lunn (2014), Should we build more large dams? The actual costs of hydropower megaproject development, *Energy Policy*, *69*, 43–56, doi:10.1016/j.enpol.2013.10.069.

Bellman, R. (1957), *Dynamic Programming*, Princeton Univ. Press, Princeton, N. J.

Beltrame, L., D. Carbonin, S. Galelli, A. Castelletti, and M. Giuliani (2014), Quantifying ENSO impacts at the basin scale using the Iterative Input variable Selection algorithm, in *Proceedings of the 7th International Congress on Environmental Modeling and Software (iEMSs 2014)*, June 15–19, edited by D. P. Ames, N. W. T. Quinn, and A. E. Rizzoli, San Diego, Calif.

Bertsekas, D. (1976), *Dynamic Programming and Stochastic Control*, Academic, N. Y.

Block, P., and L. Goddard (2012), Statistical and dynamical climate predictions to guide water resources in ethiopia, *J. Water Resour. Plann. Manage.*, *138*(3), 287–298.

Bras, R., R. Buchanan, and K. Curry (1983), Real time adaptive closed loop control of reservoirs with the High Aswan Dam as a case study, *Water Resour. Res.*, *19*(1), 33–52.

Busoniu, L., D. Ernst, B. De Schutter, and R. Babuska (2011), Cross–entropy optimization of control policies with adaptive basis functions, *IEEE Trans. Syst. Man Cybern. Part B*, *41*(1), 196–209, doi:10.1109/TSMCB.2010.2050586.

Butler, D. (2007), Earth monitoring: The planetary panopticon, *Nature*, *450*, 778–781.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2008a), Receding horizon control for water resources management, *Appl. Math Comput.*, *204*(2), 621–631.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2008b), Water reservoir control under economic, social and environmental constraints, *Automatica*, *44*(6), 1595–1607.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2010), Tree-based reinforcement learning for optimal water reservoir operation, *Water Resour. Res.*, *46*, W09507, doi:10.1029/2009WR008898.

Castelletti, A., F. Pianosi, X. Quach, and R. Soncini-Sessa (2012), Assessing water reservoirs management and development in Northern Vietnam, *Hydrol. Earth Syst. Sci.*, *16*(1), 189–199, doi:10.5194/hess-16-189-2012.

Castelletti, A., F. Pianosi, and M. Restelli (2013), A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run, *Water Resour. Res.*, *49*, 3476–3486, doi:10.1002/wrcr.20295.

Côté, P., D. Haguma, R. Leconte, and S. Krau (2011), Stochastic optimisation of Hydro-Quebec hydropower installations: A statistical comparison between SDP and SSDP methods, *Can. J. Civ. Eng.*, *38*(12), 1427–1434.

Deisenroth, M., G. Neumann, and J. Peters (2011), A Survey on Policy Search for Robotics, in *Foundations and Trends in Robotics*, vol.2, pp. 1–142. [Available at http://www.complacs.org/uploads/CompLACS/D312.pdf.]

Delquié, P. (2008), Valuing information and options: An experimental study, *J. Behav. Decis. Making*, *21*(1), 91–109.

Desreumaux, Q., P. Côté, and R. Leconte (2014), Role of hydrologic information in stochastic dynamic programming: A case study of the Kemano hydropower system in British Columbia, *Can. J. Civ. Eng.*, *41*(9), 839–844.

Faber, B., and J. Stedinger (2001), Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts, *J. Hydrol.*, *249*(1), 113–133, doi:10.1016/S0022-1694(01)00419-X.

Fernandez, A., S. Blumsack, and P.M. Reed (2013), Operational constraints and hydrologic variability limit hydropower in supporting wind integration, *Environ. Res. Lett.*, *8*(2), 24–37, doi:10.1088/1748-9326/8/2/024037.

Fraternali, P., A. Castelletti, R. Soncini-Sessa, C. Vaca Ruiz, and A. Rizzoli (2012), Putting humans in the loop: Social computing for Water Resources Management, *Environ. Modell. Software*, *37*, 68–77.

Galelli, S., and A. Castelletti (2013a), Tree-based iterative input variable selection for hydrological modeling, *Water Resour. Res.*, *49*, 4295–4310, doi:10.1002/wrcr.20339.

Galelli, S., and A. Castelletti (2013b), Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling, *Hydrol. Earth Syst. Sci. Discuss.*, *10*(2), 1617–1655.

Galelli, S., G. Humphrey, H. Maier, A. Castelletti, G. Dandy, and M. Gibbs (2014), An evaluation framework for input variable selection algorithms for environmental data-driven models, *Environ. Modell.Software*, *62*, 33–51.

Gass, S., and T. Saaty (1955), Parametric objective function: Part II, *Oper. Res.*, *3*, 316–319.

Geurts, P., D. Ernst, and L. Wehenkel (2006), Extremely randomized trees, *Mach. Learn.*, *63*(1), 3–42.

Giuliani, M., and A. Castelletti (2013), Assessing the value of cooperation and information exchange in large water resources systems by agent–based optimization, *Water Resour. Res.*, *49*, 3912–3926, doi:10.1002/wrcr.20287.

Giuliani, M., J. Herman, A. Castelletti, and P. Reed (2014a), Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management, *Water Resour. Res.*, *50*, 3355–3377, doi:10.1002/2013WR014700

Giuliani, M., S. Galelli, and R. Soncini-Sessa (2014b), A dimensionality reduction approach for Many-Objective Markov Decision Processes: Application to a water reservoir operation problem, *Environ. Model. Software*, *57*, 101–114, doi:10.1016/j.envsoft.2014.02.011.

Giuliani, M., E. Mason, A. Castelletti, F. Pianosi, and R. Soncini-Sessa (2014c), Universal approximators for direct policy search in multi-purpose water reservoir management: A comparative analysis, in *Proceedings of the 19th IFAC World Congress*, International Federation of Automatic Control, Cape Town, South Africa.

Giuliani, M., A. Castelletti, F. Pianosi, E. Mason, and P. Reed (2015), Curses, tradeoffs, and scalable management: Advancing evolutionary multi-objective direct policy search to improve water reservoir operations, *J. Water Resour. Plann. Manage.*, doi:10.1061/(ASCE)WR.1943-5452.0000570. [Available at http://ascelibrary.org/doi/abs/10.1061/(ASCE)WR.1943-5452.0000570.]

Gleick, P., and M. Palaniappan (2010), Peak water limits to freshwater withdrawal and use, *Proc. Natl. Acad. Sci. U. S. A.*, *107*(25), 11,155–11,162, doi:10.1073/pnas.1004812107.

Graham, N., and K. Georgakakos (2010), Toward understanding the value of climate information for multiobjective reservoir management under present and future climate and demand scenarios, *J. Appl. Meteorol. Climatol.*, *49*(4), 557–573, doi:10.1175/2009JAMC2135.1.

Hadka, D., and P. Reed (2012), Diagnostic assessment of search controls and failure modes in many–objective evolutionary optimization, *Evol. Comput.*, *20*(3), 423–452.

Hadka, D., and P. Reed (2013), Borg: An auto–adaptive many–objective evolutionary computing framework, *Evol. Comput.*, *21*(2), 231–259.
Hamlet, A., and D. Lettenmaier (1999), Columbia river streamflow forecasting based on enso and pdo climate signals, *J. Water Resour. Plann. Manage.*, *125*(6), 333–341.
Hamlet, A. F., and D. P. Lettenmaier (2000), Long-range climate forecasting and its use for water management in the Pacific Northwest region of North America, *J. Hydroinform.*, *2*(3), 163–182.
Hansen, J., S. Mason, L. Sun, and A. Tall (2011), Review of seasonal climate forecasting for agriculture in sub-Saharan Africa, *Exp. Agric.*, *47*(02), 205–240.
Hart, J. K., and K. Martinez (2006), Environmental Sensor Networks: A revolution in the earth system science?, *Earth Sci. Rev.*, *78*(3), 177–191.
Hejazi, M., X. Cai, and B. Ruddell (2008), The role of hydrologic information in reservoir operation-learning from historical releases, *Adv. Water Resour.*, *31*(12), 1636–1650.
Karamouz, M., and H. Vasiliadis (1992), Bayesian stochastic optimization of reservoir operation using uncertain forecasts, *Water Resour. Res.*, *28*(5), 1221–1232.
Kasprzyk, J., P. Reed, B. Kirsch, and G. Characklis (2009), Managing population and drought risks using many-objective water portfolio planning under uncertainty, *Water Resour. Res.*, *45*, W12401, doi:10.1029/2009WR008121.
Khader, A., D. Rosenberg, and M. McKee (2013), A decision tree model to estimate the value of information provided by a groundwater quality monitoring network, *Hydrol. Earth Syst. Sci.*, *17*(5), 1797–1807, doi:10.5194/hess-17-1797-2013.
Kim, Y., and R. Palmer (1997), Value of seasonal flow forecasts in Bayesian stochastic programming, *J. Water Resour. Plann. Manage.*, *123*(6), 327–335, doi:10.1061/(ASCE)0733-9496(1997)123:6(327).
Lee, J.-D., J.-B. Park, and T.-Y. Kim (2002), Estimation of the shadow prices of pollutants with production/environment inefficiency taken into account: A nonparametric directional distance function approach, *J. Environ. Manage.*, *64*(4), 365–375.
Li, W., A. Sankarasubramanian, R. Ranjithan, and E. Brill (2014), Improved regional water management utilizing climate forecasts: An interbasin transfer model with a risk management framework, *Water Resour. Res.*, *50*, 6810–6827, doi:10.1002/2013WR015248.
Loucks, D., and O. Sigvaldason (1982), Multiple-reservoir operation in North America., in *The Operation of Multiple Reservoir Systems, IIASA Collab. Proc. Ser.*, edited by Z. Kaczmarck and J. Kindler, pp. 1–103, IIASA.
Loucks, D., E. van Beek, J. Stedinger, J. Dijkman, and M. Villars (2005), *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*, UNESCO, Paris, France.
Madani, K., and J. Lund (2012), California's Sacramento–San Joaquin delta conflict: From cooperation to chicken, *J. Water Resour. Plann. Manage.*, *138*(2), 90–99, doi:10.1061/(ASCE)WR.1943-5452.0000164.
Maier, H., et al. (2014), Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions, *Environ. Modell. Software*, *62*(0), 271–299, doi:10.1016/j.envsoft.2014.09.013.
Maurer, E. P., and D. P. Lettenmaier (2004), Potential effects of long-lead hydrologic predictability on Missouri river main-stem reservoirs, *J. Clim.*, *17*(1), 174–186.
Mhaskar, H., and C. Micchelli (1992), Approximation by superposition of sigmoidal and radial basis functions, *Adv. Appl. Math.*, *13*(3), 350–373.
Minsker, B., et al. (2006), NCSA environmental cyberinfrastructure demonstration project: Creating cyberenvironments for environmental engineering and hydrological science communities, in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 161 pp., ACM, Tampa, Fla.
Oludhe, C., A. Sankarasubramanian, T. Sinha, N. Devineni, and U. Lall (2013), The role of multimodel climate forecasts in improving water and energy management over the tana river basin, kenya, *J. Appl. Meteorol. Climatol.*, *52*(11), 2460–2475.
Pianosi, F., and R. Soncini-Sessa (2009), Real-time management of a multipurpose water reservoir with a heteroscedastic inflow model, *Water Resour. Res.*, *45*, W10430, doi:10.1029/2008WR007335.
Pianosi, F., X. Quach, and R. Soncini-Sessa (2011), Artificial Neural Networks and Multi Objective Genetic Algorithms for water resources management: An application to the Hoabinh reservoir in Vietnam, in *Proceedings of the 18th IFAC World Congress*, International Federation of Automatic Control, Milan, Italy.
Pianosi, F., A. Castelletti, and M. Lovera (2012), Identification of a flow-routing model for the Red River network, in *System Identification*, vol. 16, pp. 1037–1042, International Federation of Automatic Control, Brussels, Belgium.
Piccardi, C., and R. Soncini-Sessa (1991), Stochastic dynamic programming for reservoir optimal control: Dense discretization and inflow correlation assumption made possible by parallel computing, *Water Resour. Res.*, *27*(5), 729–741.
Powell, W. (2007), *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, John Wiley, N. J.
Reed, P., D. Hadka, J. Herman, J. Kasprzyk, and J. Kollat (2013), Evolutionary Multiobjective optimization in water resources: The past, present, and future, *Adv. Water Resour.*, *51*, 438–456.
Sakalaki, M., and S. Kazi (2007), How much is information worth? Willingness to pay for expert and non-expert informational goods compared to material goods in lay economic thinking, *J. Inform. Sci.*, *33*(3), 315–325.
Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1: A strategy for system predictor identification, *J. Hydrol.*, *239*(1), 232–239.
Sharma, A., and S. Chowdhury (2011), Coping with model structural uncertainty in medium-term hydro-climatic forecasting, *Hydrol. Res.*, *42*(2-3), 113–127.
Sheer, D. (2010), Dysfunctional water management: Causes and solutions, *J. Water Resour. Plann. Manage.*, *136*(1), 1–4.
Shukla, S., N. Voisin, and D. P. Lettenmaier (2012), Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill, *Hydrol. Earth Syst. Sci. Discuss.*, *9*(2), 1827–1857, doi:10.5194/hessd-9-1827-2012.
Soncini-Sessa, R., A. Castelletti, and E. Weber (2007), *Integrated and Participatory Water Resources Management: Theory*, Elsevier, Amsterdam, Netherlands.
Stedinger, J., B. Sule, and D. Loucks (1984), Stochastic dynamic programming models for reservoir operation optimization, *Water Resour. Res.*, *20*(11), 1499–1505.
Stone, R. (2011), The legacy of the three gorges dam, *Science*, *333*(6044), 817–817, doi:10.1126/science.333.6044.817.
Tejada-Guibert, J., S. Johnson, and J. Stedinger (1995), The value of hydrologic information in stochastic dynamic programming models of a multireservoir system, *Water Resour. Res.*, *31*(10), 2571–2579.
The Economist Editorial (2011), *Welcome to the Anthropocene*.
Tsitsiklis, J., and B. Van Roy (1996), Feature-based methods for large scale dynamic programming, *Mach. Learn.*, *22*, 59–94.
U.S. Army Corps of Engineers (1977), *Reservoir System Analysis for Conservation*, *Hydrologic Engineering Methods for Water Resources Development,* Hydrol. Eng. Cent., Davis, Calif.

Voisin, N., A. F. Hamlet, L. P. Graham, D. W. Pierce, T. P. Barnett, and D. P. Lettenmaier (2006), The role of climate forecasts in western US power planning, *J. Appl. Meteorol. Climatol.*, *45*(5), 653–673.

Yokota, F., and K. Thompson (2004), Value of information analysis in environmental health risk management decisions: Past, present, and future, *Risk Anal.*, *24*(3), 635–650.

You, J.-Y., and X. Cai (2008), Determining forecast and decision horizons for reservoir operations under hedging policies, *Water Resour. Res.*, *44*, W11430, doi:10.1029/2008WR006978.

Zhao, T., and J. Zhao (2014), Joint and respective effects of long-and short-term forecast uncertainties on reservoir operations, *J. Hydrol.*, *517*, 83–94.

Zhao, T., J. Zhao, J. Lund, and D. Yang (2014), Optimal hedging rules for reservoir flood operation from forecast uncertainties, *J. Water Resour. Plann. Manage.*, *140*(12), 04014041, doi:10.1061/(ASCE)WR.1943-5452.0000432.

Zitzler, E., K. Deb, and L. Thiele (2000), Comparison of multiobjective evolutionary algorithms: Empirical results, *Evol. Comput.*, *8*(2), 173–195, doi:10.1162/106365600568202.

Zitzler, E., L. Thiele, M. Laumanns, C. Fonseca, and V. da Fonseca (2003), Performance assessment of multiobjective optimizers: An analysis and review, *IEEE Trans. Evol. Comput.*, *7*(2), 117–132.

Ziv, G., E. Baran, S. Nam, I. Rodriguez-Iturbe, and S. Levin (2012), Trading-off fish biodiversity, food security, and hydropower in the mekong river basin, *Proc. Natl. Acad. Sci. U. S. A.*, *109*(15), 5609–5614, doi:10.1073/pnas.1201423109.